DOCUMENT RESUME

ED 171 997 CE 021 873

TITLE	Instruction Manual: Career and Occupational Development Kit.
INSTITUTION	Education Commission of the States, Denver, Colo. National Assessment of Educational Progress.
SPONS AGENCY	National Center for Educational Research and Development (DHEW/OE), Washington, D.C.
PUB DATE	[71]
GRANT	0-9-08771-2468(508)
NOTE	66p.; Tables may be marginally legible
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Career Development; *Educational Assessment; *Educational Objectives; National Surveys; *Program Evaluation; *Scoring Formulas; Test Construction; *Testing; Test Interpretation; Test Items

ABSTRACT

This instruction manual was developed as a pilot by the National Assessment of Educational Progress (NAEP) to help educators conduct program and pupil evaluations and/or needs assessments in the area of career and occupational development (COD). An introduction presents the history of the National Assessment, explains the variety of uses for NAEP materials, and tells what is included in this assessment kit. The remainder of the manual is divided into four chapters. Chapter one contains the COD objectives, samples of the assessment items developed from these objectives, a description of the item format, general guidelines for using these materials, and guidelines for modifying the items and for generating other evaluation items. Chapter two provides instruction for administering the items, and chapter three provides instructions for using the results. Guidelines for scoring the items and tallying the results are presented in chapter four. The following information is appended: a discussion of survey sampling; a table of the items showing administration type, scoring type, and timings; definitions for NAEP reporting groups; and NAEP scoring rules for open-ended items. (BM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED171997

INSTRUCTION MANUAL

CAREER AND OCCUPATIONAL DEVELOPMENT KIT

CE 021 813

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

DISCRIMINATION PROHIBITED

Title VI of the Civil Rights Act of 1964 states: "No person in the United States shall, on the ground of race, color, or national origin, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance." Title IX of the Education Amendments of 1972, Public Law 92-318, states: "No person in the United States shall, on the basis of sex, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any education program or activity receiving Federal financial assistance." Therefore, career education projects supported under Sections 402 and 406 of the Education Amendments of 1974, like every program or activity receiving financial assistance from the U.S. Department of Health, Education, and Welfare, must be operated in compliance with these laws.

The material in this publication was prepared pursuant to a grant or contract from the Office of Education, U.S. Department of Health, Education, and Welfare. However, points of view or opinions expressed do not necessarily represent policies or positions of the Office of Education.

NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS
Suite 700, 1860 Lincoln
Denver, Colorado 80295

TABLE OF CONTENTS

LIST OF FIGURES, TABLES AND EXHIBITS. v

FOREWORD. vii

INTRODUCTION. ix
What Is National Assessment?. ix
Uses for These Materials. ix
What Is Included in the Assessment Kit? x

CHAPTER 1 The Assessment Objectives and Items 1
The Objectives for Career and Occupational Development 1
The Assessment Items. 2
Item Formats. 4
General Guidelines for Using These Materials. 7
Modifying the Items 7
Generating Other Evaluation Items: Sources for Program Ideas 8

CHAPTER 2 Administering the Items 13
Comparing Your Results With Assessment Results. 13

CHAPTER 3 Using the Results 19
Comparing Your Results to National Assessment Results 19
Calculating Average Performance for Individual Students 20
Other Uses of the Data. 21

CHAPTER 4 Scoring the Items and Tallying the Results. 23
Scoring Multiple-Choice Items 23
Scoring Open-Ended Exercises. 25
Tallying Responses. 27
Scoring Individually Administered Exercises 27
Possible Modifications of Scoring Procedures. 29
Possible Modifications of Exercise Tallying Procedures. 30

APPENDIX A Discussion on Survey Sampling 31
Deciding Whether Sampling Can Be of Benefit to You. 31
Overview of Steps in Sampling 32

Choosing Which Sampling Approach Will Be Most Useful for Your Purposes 34
Determining Population Estimates From Sample Data 38
Estimating the Reliability of Your Findings 39
Comparing Your Data to National Assessment Data. 41

APPENDIX B Table of Items: Showing Administration Type, Scoring Type and Timings. 51

APPENDIX C Definitions of National Assessment Reporting Groups. 53
Population Groups Included in the Study 53

APPENDIX D National Assessment Scoring Rules for Open-Ended Items. 57

LIST OF FIGURES, TABLES AND EXHIBITS

FIGURE 1.	Sample Exercise	2
FIGURE 2.	Sample Exercise Coding Line	3
TABLE 1.	Tallies of Exercises: by Age, Objective, Overlaps	4
FIGURE 3.	Example Individually Administered Exercise	5
TABLE 2.	Tallies of Items: by Age Level, Scoring Type and Administration Type	6
FIGURE 4.	Changing an Open-Ended Item to a Multiple-Choice Item	8
FIGURE 5.	Changing an Individually Administered Item to a Group-Administered, Multiple-Choice Item	9
FIGURE 6.	Using a Given Item as the Basis for Developing a New Item	10
FIGURE 7.	Item Developed to Measure an Illustrative Behavior Not Measured by Other Items	11
FIGURE 8.	Which Procedures Should You Replicate?	14
FIGURE 9.	Checklist for Replicating National Assessment Procedures	16
FIGURE 10.	Sample Tally of Results for a Multiple-Choice Item	24
FIGURE 11.	Sample Scoring of Open-Ended Item and Tallying of Results	26
FIGURE 12.	Sample Tally Sheet	27
TABLE A-1.	Table of Approximate Sample Sizes for Various Levels of Sampling Error	33
TABLE A-2.	Estimates of Standard Errors for P-Values When Sampling Fractions Are 0 to .09	43

TABLE A-3. Estimates of Standard Errors for P-Values When Sampling Fractions Are .10 to .19 44

TABLE A-4. Estimates of Standard Errors for P-Values When Sampling Fractions Are .20 to .29 45

TABLE A-5. Estimates of Standard Errors for P-Values When Sampling Fractions Are .30 to .39 46

TABLE A-6. Estimates of Standard Errors for P-Values When Sampling Fractions Are .40 to .49 47

TABLE A-7. Estimates of Standard Errors for P-Values When Sampling Fractions Are .50 to .59 48

TABLE A-8. Estimates of Standard Errors for P-Values When Sampling Fractions Are .60 to .69 49

TABLE A-9. Estimates of Standard Errors for P-Values When Sampling Fractions Are .70 to .79 50

EXHIBIT 1. National Assessment Geographic Regions. 54

FOREWORD

This career and occupational development kit is a pilot effort developed by the National Assessment of Educational Progress under a grant from the Office of Career Education, U.S. Office of Education. We hope that these materials will be useful to you in conducting program and pupil evaluations and/or needs assessments. It is difficult to define procedures that will apply to everyone, but we hope that these instructions will be sufficient for you to proceed on your own.

We ask that you note any problems you encounter in using the materials and let us know about them. Your candid evaluation of this kit's strengths and weaknesses will help us decide on development of future assessment kits.

INTRODUCTION

Although many of the concepts involved in career education are not new, the emergence of career education as a separate reform effort has been fairly recent. At present, the testing instruments to measure achievement and program effectiveness in this field are limited. To assist those faced with evaluating or assessing the need for career-education programs, the National Assessment of Educational Progress (NAEP), under a grant from the Office of Career Education, U.S. Office of Education, is making available some of the materials used in its assessment of national performance in career and occupational development (COD).

What Is National Assessment?

The National Assessment is a national survey of knowledge, skills, understandings and attitudes of young Americans aged 9, 13 and 17 and young adults aged 26-35 in major learning areas, funded by the National Center for Education Statistics. The surveys in each learning area are repeated periodically to monitor changes in achievement levels over time.

National Assessment conducts objectives-referenced assessments -- in other words, its test items are developed to measure achievement of stated objectives. Results are reported for groups of students, not for individual students.

The Assessment makes its findings available to education decision makers and to all those interested in the field of education. It also provides materials to state and local education agencies at their request and assists them in using the materials if they so desire.

Uses for These Materials

This kit brings together different types of items that measure a number of career and occupational development objectives. These materials can be used in a variety of ways.

-- Results for a particular classroom, school or district might be used to make general comparisons with NAEP national age-level results. If the particular school or district has large proportions of any of National Assessment's reporting groups -- for example, blacks, advantaged urban residents, extreme-rural residents or people whose parents have low levels of education -- it may compare local results with NAEP results for these groups.
-- These items can also be used as a diagnostic instrument. Results could be obtained either for individuals or groups of students (a classroom,

a grade level, a school or a district). Percentages correct could also be calculated for different groups within the school population -- for example, those taking a certain type of academic program, those participating in federally-assisted programs or those from different socioeconomic-status groupings.

-- These materials may be used as a basis for generating a tailored career-education evaluation instrument for one's own situation. Items designed to be administered to individuals could be adapted for classroom groups; items administered at one age might be modified to apply to other age levels. The questions also might serve as a stimulus for developing other items using either a similar format or similar content material.

-- The objectives and items might also be used as suggestions for content or activities for career-education programs.

-- Various combinations of the above approaches might be used.

These materials are not intended to provide a complete evaluation of any career-education program. National Assessment feels strongly that no one set of items is adequate to "define" an area as broad as career education. NAEP's items measure objectives developed by National Assessment (see beginning of Chapter 1). These objectives may or may not correspond, in whole or in part, to the career-education objectives of other programs. Also, the items included here are only a part of the complete set of items used by NAEP in its assessment. Some items are being reserved to measure changes in achievement over time; some items concerned with basic skills are not included, since they are items of a type readily available from other sources.

Some of the items contain job titles or pronoun references that were not modified in accord with current standards against sex stereotyping. These items can be changed if you wish, although results then will not be strictly comparable to those of National Assessment.

What Is Included in the Assessment Kit?

The kit includes the following materials:

-- Objectives for career and occupational development: These objectives, published in 1971, represent a consensus view of subject-matter consultants and lay persons about what was important for students to achieve in career and occupational development. Assessment items were developed to measure achievement of these objectives.

-- Assessment items: The items presented here are a part of the total COD assessment. They include tasks measuring self-appraisal skills, work-related experiences, knowledge about jobs and attitudes towards work.

-- Instruction manual: This manual (which you are now reading) provides information on how to use the materials, how to administer the items and how to interpret the results.

-- Scoring guides and national percentages of response: These guides show correct answers for every item included. In the case of open-ended

items, categories of responses, with examples of responses scored in each category, are given. Included with the scoring guides are national percentages of response for each item for the ages it assessed.

CHAPTER 1

THE ASSESSMENT OBJECTIVES AND ITEMS

The assessment items are designed to measure the achievement of career and occupational development objectives (COD) by different age groups of students. Each assessment is the product of several years' work by a great many educators, scholars and lay persons from all over the country. Initially, these people design objectives for a learning area, proposing specific goals that they feel Americans should be achieving in the course of their education. These goals are reviewed by more people and then given to the test developers, whose task it is to create measurement tools appropriate to the objectives.

When the items prepared by the test developers have passed extensive reviews by educators, scholars and lay persons, they are administered to a probability sample of individuals at each age level (9, 13, 17 and young adult). The people who compose the sample are chosen in such a way that the results of their assessment can be generalized to the entire population at that age.

The Objectives for Career and Occupational Development

Before assessing what children know, it is necessary to determine what you think they should know. This is one function of the objectives: to set goals against which to measure performance. Following are the major objectives and subobjectives that formed the basis for the COD assessment.

Summary of Career and Occupational Development Objectives

I. Prepare for making career decisions
   - A. Know own characteristics relevant to career decisions
   - B. Know the characteristics and requirements of different careers and occupations
   - C. Relate own personal characteristics to occupational requirements
   - D. Plan for career development or change

II. Improve career and occupational capabilities

III. Possess skills that are generally useful in the world of work
   - A. Have generally useful numerical skills
   - B. Have generally useful communication skills
   - C. Have generally useful manual/perceptual skills

D. Have generally useful information-processing and decision-making skills
E. Have generally useful interpersonal skills
F. Have employment-seeking skills

IV. Practice effective work habits
A. Assume responsibility for own behavior
B. Plan work
C. Use initiative and ingenuity to fulfill responsibilities
D. Adapt to varied conditions
E. Maintain good health and grooming

V. Have positive attitudes toward work
A. Recognize the bases of various attitudes toward work
B. Hold competence and excellence in high regard
C. Seek personal fulfillment through own achievements
D. Value work in terms of societal goals

The order of the objectives is not necessarily an indication of their relative importance. Each item is coded to indicate the objective and subobjective that it measures. A copy of the complete career and occupational development objectives, including specific behavioral examples for each subobjective, is a part of this kit.

The Assessment Items

The assessment items are printed as they appeared in the National Assessment test booklets. They are included in loose-leaf form so that they can be easily sorted, rearranged, reviewed and duplicated. A sample item appears in Figure 1.

FIGURE 1. Sample Exercise

A. People can do some things better than other things. What are two things that you can do well?

(1) _____
_____
(2) _____
_____

B. What are two things you have tried that you do NOT do very well?

(1) _____
_____
(2) _____
_____

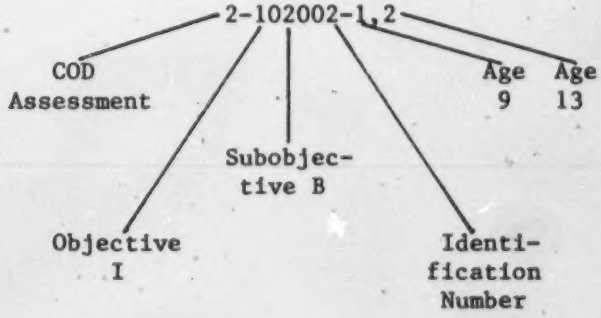
2-101003-1d,2

Each item may contain more than one part and may take up more than one printed page. Each item is identified by a unique coding number printed in the lower left-hand corner of each page. Item parts are identified by capital letters -- A, B, C and so forth. The item shown in Figure 1 has two parts, part A and part B.

Interpreting the Coding Line

The unique coding number (lower left-hand corner of each page) indicates the objective and subobjective measured by an item and the age group(s) for which it is intended. Figure 2 shows a sample exercise coding line. The first number -- always a 2 -- means that the item is part of the career and occupational development assessment. The six digit number gives the objective and subobjective being assessed: the first number indicates the major objective (1 = I, 2 = II and so on), and the second and third numbers show the subobjective (01 = A, 02 = B and so forth). The last three digits are a unique number for exercise identification purposes, but have no other significance. The numbers following the second dash indicate age levels at which the exercise is administered (1 = 9, 2 = 13, 3 = 17, 4 = adult). In some cases a small "d" follows an age level number; this indicates that the exercise varied slightly in format for the different ages. A "31" means that the exercise was only administered to 17-year-olds who were attending school at the time of the assessment.

FIGURE 2. Sample Exercise Coding Line



Numbers of Exercises

Thirty-six unique items are included in this kit. The number of parts (A, B, C and so forth) in an item varies with each item. In some cases, parts of an item could be administered independently; thus, the total number of items available depends somewhat on how they are used.

Some items are designed to be administered to more than one age group; these items are termed "overlap" exercises by National Assessment. Comparisons

across age groups should only be made for identical (overlap) exercises or identical groups of exercises. Table 1 shows objective-by-objective and total numbers of exercises for each age level and the numbers of identical (overlap) exercises for various combinations of age levels.

TABLE 1. Tallies of Exercises:
by Age, Objective, Overlaps

Total Number of Exercises at Each Age, by Objective

Objective	Total	Age 9	Age 13	Age 17
I	15	4	10	12*
II	4	3	3	2
III	9	3	7	8
IV	5	3	4	3
V	5	2	4	4
TOTAL	38	15	28	29

Total Number of Overlap Exercises at Various Ages, by Objective†

Objective	Ages 9-13	Ages 13-17	Ages 9-13-17**
I	3	8	1
II	3	1	1
III	3	6	2
IV	3	2	1
V	2	3	1
TOTAL	14	20	6

**Numbers do not add across due to overlap exercises.*
†Items with different formats (group v. individual) but identical content are considered overlap exercises.
***9-13-17 overlaps are also included in 9-13 and 13-17 counts so that complete information is available for comparisons of two age levels.*

Item Formats

Three types of item formats are included in this kit: (1) multiple-choice items designed to be administered to groups of students, (2) open-ended items to be administered to groups of students and (3) items administered in an interview situation to one student at a time. Individually administered questions are generally open-ended.

The multiple-choice items were used in situations in which recognition of the correct answer was sufficient fulfillment of the objective. Open-ended questions measured concepts that were difficult to cast in multiple-choice format and objectives that required students to generate answers rather than recognize them. The "individual" exercises are identified by the box marked "Comments" on the first page of the exercise (see Figure 3). The boxes were for the use of the National Assessment test administrators and may be disregarded. These items were most often administered individually because they involved contingent questions -- that is, the answer to one question would determine the next question to be asked. Some 9-year-old items were individually administered because the writing competence required was felt to be above 9-year-old abilities. Nine-year-olds answered aloud and the test administrator wrote down what they said. Table 2 shows the number of items of various formats for each age level.

FIGURE 3. Example Individually Administered Exercise

Comments:

_____

_____

_____

A. There are tests to help you decide on your career or job plans. These tests help you to find out such things as what you like to do best, what jobs you would rather have, or if you have mechanical, clerical or other abilities. Did you ever take such a test?

- ⬭ Yes (Go to B)
- ⬭ No (End the exercise)
- ⬭ I don't know. (End the exercise)
- ⬭ No response (End the exercise)

B. Did you discuss the test results with anyone who could advise you on your plans?

- ⬭ Yes (Go to C)
- ⬭ No (End the exercise)
- ⬭ I don't know. (End the exercise)
- ⬭ No response (End the exercise)

TABLE 2. Tallies of Items: by Age Level, Scoring Type and Administration Type

	Multiple-Choice	Open-Ended	Totals
Age 9			
Administered to group	5	6	11
Administered individually	0	4	4
TOTALS	5	10	15
Age 13			
Administered to group	8	14	22
Administered individually	0	6	6
TOTALS	8	20	28
Age 17			
Administered to group	5	16	21
Administered individually	0	8	8
TOTALS	5	24	29

All items differ from those found in standardized tests in that students are asked to mark or write their answers directly on the item instead of using a separate answer sheet. All of the multiple-choice exercises include "I don't know" as one of the answer alternatives. National Assessment uses this alternative as an attempt to reduce student guessing.

These item formats were designed for the National Assessment's administration procedures and may require some modifications for other uses. For example, the open-ended items, while no more difficult to administer than multiple-choice questions, are more difficult to score. Adminstering items to individual students one at a time may create logistical difficulties. You may wish to change open-ended items to multiple-choice items or individually administered items to a form suitable for groups of students. You may also wish to design separate answer sheets to facilitate scoring and to lessen the amount of duplicating needed. If you intend to compare your results to National Assessment's, the amount of modifications that you can make will be limited. However, if you do not intend to make such comparisons, many types of modification are possible.

Found in the next chapters are specific instructions for using these materials to compare your results to NAEP's and suggestions for modifications in procedures if you do not wish to make comparisons with NAEP. Following are some general guidelines which should be attended to before making any use of the items.

General Guidelines for Using These Materials

The first step in any use of this kit should be a careful examination of all the materials. This should include a review of the objectives to determine which ones are compatible with your goals for your students. After deciding which objectives are relevant to your needs, you should examine the items measuring those objectives and become familiar with the particular facet of an objective or subobjective assessed by each item. You should determine whether you are in agreement with the scoring keys and scoring guides for the items, and use this information in deciding which items will be appropriate to your purposes.

The next step is to decide how these materials can be most useful in evaluating your programs and your students' abilities or in assessing student needs. Some options are: to compare your results with those obtained by National Assessment, to construct your own test using some or all of the NAEP items, to modify items before using them, to use items as a starting point for generating other items or to use the objectives as a source either of program suggestions or ideas for evaluation items.

Modifying the Items

You may find that the format of the items makes them difficult to use in a classroom or school setting. You may wish to change the language or job titles in some items to meet current standards against sex stereotyping. If comparison with National Assessment results is not of primary importance, it may be more convenient to modify some items.

For instance, you may wish to devise a separate answer sheet. For the multiple-choice items, alternatives should be numbered consecutively from top to bottom and students could either darken a numbered space or write down the number of the correct answer. Spaces might be left so students could write their answers to open-ended questions on the answer sheet.

The open-ended questions are, in some cases, quite time-consuming to score. One alternative might be to convert them to multiple-choice items. Categories listed as incorrect responses in the item scoring guide might be used as a source for answer alternatives. Percentages listed on the scoring guides indicate responses most frequently given. Figure 4 shows an example of this type of change. If items are still to be open-ended but wording is changed substantially, scoring guides may need to be modified.

The individually administered items will probably require some modifications. The teacher might choose to read the items aloud and have the students write their answers on a separate answer sheet. Or more explicit instructions for contingent questions might be included on the item and the boxes, which were for the use of NAEP administrators, blanked out. As with the open-ended items above, some of these questions could be converted to multiple-choice. Figure 5 shows an example of these types of conversions.

FIGURE 4. Changing an Open-Ended Item to a Multiple-Choice Item

Original Open-Ended Item and Scoring Guide

Suppose a place where you worked made a new rule you thought was very unfair. What would you do?

1. _____

2. _____

Scoring Guide and Results for Open-Ended Item

Age 17 -- At Least One Response in Category	Scoring Categories
Acceptable categories	
18.9%	Protest or petition
70.2	Discuss with someone in authority
20.0	Discuss with fellow workers
Unacceptable categories	
3.7	Complain about it
12.7	Passive obedience
0.7	Ignore the rule
19.3	Quit and find other employment

Multiple-Choice Version

Suppose a place where you worked made a new rule that you thought was very unfair. What would you do?

- ⬭ Complain to my friends about it.
- ⬭ Obey the new rule.
- ● Talk about it with my supervisor.
- ⬭ Quit and look for another job.
- ⬭ I don't know.

Generating Other Evaluation Items: Sources for Program Ideas

Both the items and the objectives can be used as idea sources for generating other items. One easy example would be the items measuring knowledge

FIGURE 5. Changing an Individually Administered Item to a Group-Administered, Multiple-Choice Item

This is an adaptation of Item 101016. Please see that item for the original version.

Modified Item 101016

A. There are tests to help you decide on your career or job plans. These tests help you to find out such things as what you like to do best, what jobs you would rather have, or if you have mechanical, clerical or other abilities. Did you ever take such a test?

⬭ Yes

⬭ No

⬭ I don't know.

(If you answered yes, go on to part B.)

B. Did you discuss the test results with anyone who could advise you on your plans?

⬭ Yes

⬭ No

⬭ I don't know.

(If you answered yes, go on to part C.)

C. What job or position did this person have?

⬭ School counselor

⬭ Employment counselor

⬭ Psychiatrist

⬭ Teacher

⬭ Military counselor

⬭ Friend(s)

⬭ Parents

⬭ Other

⬭ I don't know.

about jobs -- other items using the same format with slightly different content would not be difficult to develop (see Figure 6). The "illustrative behaviors" listed under COD subobjectives are not all covered by the items given here; items can be developed for behaviors that have not been measured as shown in Figure 7.

FIGURE 6. Using a Given Item as the Basis for Developing a New Item

Original Item	New Item
Which person usually earns the most money?	Which person usually earns the most money?
○ Elementary school teacher	● Systems analyst
○ Postal clerk	○ Computer programmer
○ Nurse	○ Keypunch operator
● Chemist	○ Computer operator
○ I don't know.	○ I don't know.

Items and objectives can also be used as a source of program ideas. For example, many of the individually administered items could serve as the basis for class discussions. The objectives provide many examples of behaviors that might be included in the curriculum.

FIGURE 7. Item Developed to Measure an Illustrative Behavior Not Measured by Other Items

Objective I. Prepare for making career decisions

B. Know the characteristics and requirements of different careers and occupations

2. Know differences in work conditions among and within occupational families

Illustrative Behavior

Ages 13, 17 and Adult -- Recognize that occupations influence the kind of home life persons have. (This behavior is found on page 19 of the COD objectives booklet.)

A. Does a policeman frequently have to work at night?

- ● Yes
- ○ No
- ○ I don't know.

B. Does a bank officer frequently have to work at night?

- ○ Yes
- ● No
- ○ I don't know.

C. Does an airplane pilot frequently have to work at night?

- ● Yes
- ○ No
- ○ I don't know.

D. Does a judge frequently have to work at night?

- ○ Yes
- ● No
- ○ I don't know.

E. Does a long distance truck driver frequently have to work at night?

- ● Yes
- ○ No
- ○ I don't know.

CHAPTER 2

ADMINISTERING THE ITEMS

In making these materials available, National Assessment wished to make it as convenient as possible for people to experiment with various ways of using the materials. These may range from duplicating the NAEP assessment as closely as possible and comparing results to NAEP's to using National Assessment objectives for program ideas and bypassing the items entirely for the time being. This chapter first provides instructions for repeating NAEP administration procedures and then gives suggestions for modifications of these procedures.

Comparing Your Results With Assessment Results

The more closely that you follow National Assessment procedures, the more reliable will be the comparisons you make with NAEP results. Thus, the instructions for this section are fairly detailed. You must decide which procedures are important and possible for you to repeat and which are not. See Figure 8 for things to consider in making these decisions. If you do change some of the procedures -- for example, testing at a different time of year or testing grade levels rather than age levels -- these differences should be noted when you report comparison data.

Selecting Students To Be Assessed and Scheduling the Testing

National Assessment selects students to be assessed by age rather than grade in school. Students of ages 9, 13 and 17 are tested. For the purposes of the assessment, 9- and 13-year-olds are defined as any student born from January 1 to December 31 of the year 9 or 13 years prior to the beginning year of the assessment. Seventeen-year-olds are defined as those born from October 1 to September 30 in the years 17 years prior to the assessment years. For example, to be included in a 1977-78 school year assessment, a 9-year-old would have been born in 1968, a 13-year-old in 1964 and a 17-year-old from October 1, 1960, to September 30, 1961. The following three groups were excluded from the National Assessment sample, since the instruments were not designed for administration to these groups: educable mentally retarded students, functionally disabled students and non-English-speaking students. Although National Assessment includes both 17-year-olds who are in school and those who are not in school in the assessment, the national results given in this kit are only for 17-year-olds attending school at the time of the assessment.

FIGURE 8. Which Procedures Should You Replicate?

If you repeat the National Assessment procedures given here, you should be able to make fairly reliable comparisons with NAEP data. However, some of these procedures are costly and can be modified without a great drop in comparability.

1. How are your results to be used?

   If the comparisons made will, by themselves, have a large impact on decision making, then it is important that the comparisons be reliable. If the comparisons will be used as one piece of data, in conjunction with other information, then more latitude can be allowed.

2. Which procedures can be changed without a great drop in reliability?

   a. Paced tapes are costly, and much the same effect can be obtained by reading the materials aloud.
   b. Using separate item booklets for each student is also expensive and may cause scoring difficulties. Using separate answer sheets should not affect comparisons greatly.
   c. Results will probably be more affected by changing the time of year of testing or the age groups being assessed, since this will change the amount of schooling that the respondents have had prior to testing.
   d. If possible, time allowed for each item should be close to that given by the National Assessment.

Changes in procedures will not prevent comparisons, but any such changes should be noted in publications comparing results, so that readers can evaluate for themselves their potential effect on the data.

If you are assessing performance of a large group of students, you may wish to assess only a sample of the students rather than the entire population.[1] NAEP assesses 9-year-olds in January-February, 13-year-olds in October-December and 17-year-olds in March-April. So that your students at these age levels have been exposed to comparable amounts of schooling as the assessment sample, tests should be given to these age levels at as close to the same time of the school year as possible.

---

[1]A discussion of sampling and estimating reliability of results is found in Appendix A.

Preparing the Assessment Items

You will have already chosen items to administer that measure objectives you feel your students should accomplish. A complete set of the items to be administered, including the sample exercise, should be duplicated for each student participating in the test. If you are only testing groups of students, you will omit the individually administered exercises (identified by the "Comments" box on the first page).

Administering the Assessment Items

To minimize the possible effect of student reading abilities upon assessment performance, National Assessment used a paced tape that read questions to the students as they followed along in their test booklets. This also insured that all students had the same amount of time to work on a particular item. To keep costs reasonable, paced tapes were not provided with this kit. However, much the same effect can be obtained by having the teacher or test administrator read the materials aloud or by recording your own tape. The test administrator should also indicate when students are to stop and when they are to turn pages.

Students should also be given about the same length of time to answer a question that students in the national sample had. A list of time allowed for each of the exercises administered in a group situation is given in Appendix B. Students were allowed as much time as they needed for each part of the "individual" exercises.

Testing Different Groups of Students

If you find that grade-level rather than age-level data would be more useful, National Assessment does have available grade-level data for the age levels assessed (9-, 13- and 17-year-olds). These data were too extensive to include here; if you are interested in obtaining them, please order the Career and Occupational Development Technical Report: Exercise Volume, Report 05-COD-20 from the Superintendent of Documents (about $10). To compare your students to the grade-level data, determine which of your students in the 4th, 8th and/or 11th grades also fit the age-level criteria. Calculate scores for these students only and compare their percentages of success to the NAEP percentages for the given grade levels.

National Assessment also reports results for other population groups within the given age groups. Data are available for sex, race, level of parental education, size and type of community, grade level and home environment factors; these data are also provided in the Career and Occupational Development Technical Report: Exercise Volume. Definitions of the reporting groups used are presented in Appendix C.

Figure 9 shows a checklist of things to be considered when replicating National Assessment procedures.

FIGURE 9. Checklist for Replicating
National Assessment Procedures

1. Have you chosen items that are in accord with your objectives for your students?

2. Do you agree with the scoring guides for the items you have chosen?

3. Have you determined what groups will be assessed: 9-, 13- and/or 17-year-olds? Grades 8 or 11? Other?

4. Have you excluded educable mentally retarded, physically handicapped and non-English-speaking students?

5. Have you determined when during the school year you will administer the assessment? On the same time schedule as NAEP or other?

6. Have you decided whether you will assess all students in the group of interest or will assess a sample?

7. Have you duplicated items for each student or provided items that can be reused and an answer sheet for each student?

8. Have you determined whether you will use a tape recording, read materials aloud or let students read for themselves?

9. Have you made arrangements to regulate the time allowed for students to respond to each exercise, in accord with the time allowed by NAEP (shown in Appendix B)?

10. Have you set aside space and class time for administering the assessment?

11. Have you made provisions for scoring the items and tallying the results?

Modifying the Assessment Procedures

You may wish to use some or all of these items to make up your own individualized evaluation instrument. In making up your own test, the only constraints are those dictated by your intended use of the results. You would not have to conduct your testing at the same time of year as the Assessment and you could test any ages and/or grade levels that you desired. Reading the items to the students would depend upon whether you felt that the reading ability would interfere with their ability to answer and whether or not you felt that reading ability should be a factor in ability to answer the questions. Time given to answer questions could be left to the discretion of the test

administrator, or, if doing more than one classroom, the discretion of those responsible for test administration. The number of open-ended exercises you use will probably be determined by the scope of testing you are doing and your capabilities for scoring. Your use of individual items should be determined by your need for the type of information they elicit.

You may wish to make up your own test by modifying some or all of the items, as discussed in Chapter 1.

CHAPTER 3

USING THE RESULTS

National Assessment collects survey data; its function is to report upon the existing status of American education. The Assessment does not set standards for what should be happening in education -- it merely describes what is happening. This chapter suggests a number of different ways to summarize data and to establish standards for comparisons. Methods of scoring items and tallying data for each item are described in Chapter 4.

Comparing Your Results to National Assessment Results

One method of evaluating your results, discussed extensively in this manual, is to compare your results to those obtained by National Assessment. The NAEP results given here represent the average national performance on each item. Although this is a somewhat arbitrary standard, it is an objective measure that takes into account the abilities of many diverse groups. Results can be compared to NAEP results on an exercise-by-exercise basis. The percentage correct for each exercise or exercise part is calculated using the following formula:

$$\frac{\text{Number of students responding correctly}}{\text{Total number of students responding}} \times 100$$

Some statistical cautions are in order before you begin comparing your results to National Assessment's. The percentages given here are not absolutely exact figures, since some errors in test data are unavoidable. Errors can be divided into two major categories -- sampling and nonsampling errors. Since NAEP assesses a sample of people rather than the entire population at an age level, results are an estimate of the percentage of the entire population and, thus, are subject to sampling error.[1] Nonsampling errors are caused by such things as variations in administration and in scoring procedures. It is difficult to control these variations and it is also difficult to calculate precisely their effects. Apparent differences in performance thus may be due to

---

[1]Standard errors for National Assessment results are found in the Career and Occupational Development Technical Report: Exercise Volume, Report 05-COD-20, available from the Superintendent of Documents (about $10).

external factors rather than actual differences in ability. Although the National Assessment results provide a standard against which to measure your results, you will have to determine how much of a difference is cause for concern.

You may decide that comparisons with the national performance levels are not meaningful if your class or school has a population very different from a national "average." Definitions of the subgroups for which National Assessment reports results are given in Appendix C. You may wish to compare your results on exercises to results for any one of these subgroups.[2] To make comparisons of different types of students, students in the class or school might be placed in groups according to sex, race, socioeconomic status and so forth, and the average performance level on an exercise for a group calculated. Sampling and nonsampling errors should again be considered when evaluating magnitude and importance of differences in performance.[3]

Exercise-by-exercise results may prove to be too much data for making comparisons. You may wish to aggregate the data over a number of items and calculate the average performance. This can be done with National Assessment data as well as with your own data and comparisons could then be made. A similar procedure could be used to make comparisons with subgroup data (sex, race, etc.).

Calculating Average Performance for Individual Students

You may wish to compute total scores for individual students even though these scores would not be comparable to NAEP data, since National Assessment does not compute individual scores. A tally sheet should be kept for each student and a percentage score calculated using the following formula:

$$\frac{\text{Number items/or item parts correct}}{\text{Total number items/or item parts}} \times 100$$

You can compute an average for the entire test or you can average results for subgroupings of items, for example, items measuring a single objective within the test. The number of items or item parts in a subgrouping should be large enough to provide what you feel is a reliable average. If the purpose of the test is to furnish diagnostic information about individual students, the scoring tally forms will have to indicate not only how many items the student answered correctly, but also which items the student answered correctly and incorrectly.

---

[2]Results and standard errors for population subgroups are found in the Career and Occupational Development Technical Report: Exercise Volume.
[3]Ibid.

Other Uses of the Data

Results can also be used for ranking individuals, classrooms or schools. Again, care should be taken not to put undue emphasis on small differences in performance, since some variation due to sampling or measurement error is to be expected.

After you have looked at the results and have a general feel for your students' performance, you may wish to identify specific areas of strength and weakness. One logical grouping to consider would be performance on specific National Assessment objectives. Results from specific exercises might also indicate program areas needing attention.

CHAPTER 4

SCORING THE ITEMS AND TALLYING THE RESULTS

Scoring and tallying procedures are linked together, since the type of results desired will affect the extent of scoring done. Both must be planned together and the information to be obtained specified before scoring begins. If you wish to compare your results to National Assessment's, you will need to follow their scoring procedures and specifications as closely as possible. If you are gathering data for other uses, you can make modifications appropriate to your own situation. For any scoring system, however, the scoring keys and rules for scoring should be sufficiently clear and detailed to ensure that each student's response is scored consistently. The system of recording and tallying scores should be simple and clear to minimize clerical and transcribing errors. All possible efforts should be made to avoid scorer bias.

Before beginning to score the items, each item and scoring key should be carefully reviewed. Then define the information to be obtained from each item. So that each response only has to be scored once, it is important to determine all information wanted and instruct scorers in recording this information before they begin scoring.

The National Assessment items are designed so that responses are hand-coded in ovals at the bottom of the page and then tallied by an optical scanning machine. Using the machine-readable ovals to code responses would be too confusing for manual tallying. Thus, you will need to designate places on the exercises to record various types of information. You will also need to design tally sheets in order to aggregate information.

Scoring Multiple-Choice Items

National Assessment scores multiple-choice items with an optical scanning machine, a process that will probably be impossible for you to duplicate. Any method that you normally use for scoring multiple-choice items can be used. Correct answers for multiple-choice items are shown in the scoring guide booklet.

For multiple-choice items with several parts, you may wish to report the percentage correct on each item part and you may also wish to report the number of parts correct. To report number of parts correct, scorers would be instructed to mark this information on the exercise. An example of the way that an item might be marked is shown in Figure 10.

FIGURE 10. Sample Tally of Results for a Multiple-Choice Item

Number Correct

A. Which one of the following jobs usually requires the LONGEST period of training?

/

- ○ Manicurist
- ● Plumber
- ○ Assembly line worker
- ○ Truck driver
- ○ I don't know.

B. Which one of the following jobs usually requires the LONGEST period of training?

- ○ Lawyer
- ○ Computer operator
- ● Electronic technician
- ○ Insurance salesman
- ○ I don't know.

C. Which one of the following jobs usually requires the LONGEST period of training?

/

- ● Registered nurse
- ○ Practical nurse
- ○ Hospital orderly
- ○ X-ray technician
- ○ I don't know.

Number Correct

D. Which one of the following jobs usually requires the LONGEST period of training?

- ○ Airlines ticket agent
- ○ Tool designer
- ○ Model
- ● Draftsman
- ○ I don't know.

E. Which one of the following jobs usually requires the LONGEST period of training?

/

- ○ Supermarket checker
- ● Civil engineer
- ○ Surveyor
- ○ Air conditioning mechanic
- ○ I don't know.

TOTAL 3

National Assessment results are shown on a part-by-part basis for the multiple-choice exercises. If NAEP has reported some combination of parts correct, this is shown in the scoring guide for the particular exercise.

Scoring Open-Ended Exercises

Scoring open-ended exercises is more of a judgmental matter than scoring multiple-choice items. For this reason, if you wish to compare your results with those of National Assessment, it is important to follow NAEP procedures and scoring guides as closely as possible.

Answers to most open-ended items are not simply scored right or wrong but are placed into categories that describe the type of response given. The categories of correct and incorrect responses, along with examples of actual answers belonging in each category, are given in the scoring guide for each exercise. An explanation of how to read the scoring guides is included with the guides.

The most essential element in open-ended scoring is consistency across all responses. For a small group of responses, such as those from a single classroom, this may not be too difficult to achieve. The teacher can go through and classify the responses, referring to the examples given in the scoring guides, and should be able to remember the entire group of responses well enough to insure consistency. The teacher should make every effort to remove anything that might allow bias to affect the scoring process. Children's names should be blocked out; if the teacher recognizes handwriting, it might be well for someone other than the classroom teacher to do the scoring.

If scoring is being done on a larger group of responses, consistency will be more difficult to achieve. The person responsible for scoring should become thoroughly familiar with scoring categories and examples and could then train others to score items. He/she would take a sample of responses for each item and score those responses. The other scorers would then score the same set of papers; scores would be compared and differences discussed. This discussion might cause some reinterpretation of the scoring system. When scorers and the trainer are in what they feel is sufficient agreement about their interpretation of the scoring guide, scorers will begin scoring other responses. The trainer should periodically check their work to insure that they continue to use the same criteria throughout.

For each response line of a particular item, the response given should be assigned a category number; a "no response" or "I don't know" answer should also be coded. You may wish to only designate responses acceptable or unacceptable, or you may wish to employ the full range of categories defined by National Assessment. An example of the way that an open-ended item might be scored is shown in Figure 11.

Students may write more than one response on a line; they may skip lines or they may write answers in an unusual order. Rules used by National Assessment for dealing with these and other unusual situations are listed in Appendix D.

FIGURE 11. Sample Scoring of Open-Ended Item and Tallying of Results

Suppose you want to find out more about a job you are interested in. What are five things you could do NOW to find out more about the job before you take the job or begin job training?

Number Acceptable	Categories Used		
1	12	(1)	Read a book about it.
1	15	(2)	Write to Company and ask them for a pamphlet.
1	11	(3)	Take a tour through the place.
	22	(4)	Take lessons.
	39	(5)	I don't know.
TOTAL 3			

Many of the open-ended items call for more than one response. The percentage correct for each line is not a particularly meaningful result by itself and other tallies will probably be desirable. To compare results with National Assessment results, consult the data sheet for a particular item and determine how NAEP summarized results for the exercise. In questions asking for more than one response, NAEP often required a certain minimum number of acceptable responses. This number will usually be shown on the scoring guides. The number of "acceptable" responses made by each student should be marked on the item or student answer sheet. (See Figure 11.) These acceptable answers can include more than one answer in the same category unless otherwise stated on the scoring guide. The number of students giving the minimum or more correct answers is then used to calculate the percentage of students performing at an acceptable level.

Another type of information that can be compiled is the categories used most often in response to an item when all response lines for an item are

considered together. These data describe the percentage of students who used a particular category but eliminates credit for mentioning the same category more than once. National percentages for those using a category at least once are listed on the scoring guides beside the category headings.

To compute this information, list (on the answer sheet or the student's item sheet) all the different categories used by a student. Determine the percentage of students who mentioned the category at least once using the following formula:

$$\frac{\text{Number of students using category at least once}}{\text{Total number of students responding to item}} \times 100$$

National Assessment used its own criteria to determine the number of "acceptable" responses to be expected. If you wish to compare your results with NAEP's, you will need to use the same number. However, if comparison of results is not a consideration, you should examine the exercises and scoring guides carefully and determine what you think would be an acceptable level for your particular students.

Several exercises have more complicated scoring criteria; complete directions for scoring these items usually appear with the scoring guides for the exercises.

Tallying Responses

Responses should be tallied as they are scored. A tally sheet could be kept for each item so that the percentage of correct responses can be calculated. Percentages for incorrect responses should be tallied if diagnostic information is desired. If scores for individual students are to be computed, a tally would also be kept for each individual. A sample tally sheet, showing tallies for multiple-choice and open-ended items, appears in Figure 12. Percentages for items and for individual students can be calculated using this sheet. This is only a sample form and you should feel free to develop your own if that proves more efficient.

Scoring Individually Administered Exercises

Each part of an individual exercise (part A, B, etc.) is scored using the scoring guide for the exercise part and the same scoring procedures as described for open-ended exercises. Before scoring is begun, determine what tallies are desired. If you wish to count the number of acceptable responses or to determine which categories were used most frequently, appropriate tallies need to be kept.

To make comparisons with National Assessment results, you will wish to use the same values that NAEP did to summarize results for a particular exercise. These are shown on the scoring guides.

Even if students are not asked some parts of individually administered exercises, due to their answers on previous parts, they should still be con-

FIGURE 12. Sample Tally Sheet

Exercises and Scoring Categories	John	Maria	Robert	Frank	Linda	Tina	Susan	Michael	Ann	Julius	Percentage for Correct Item Part
* 101002, Talked to Someone Older and Aware	/	/		/	/	/	/			/	70%
At Least One Category 10	/	/		/		/				/	50%
At Least One Category 11	/		/		/	/					40%
At Least One Category 12	/	/		/	/		/				50%
At Least One Category 21		/						/			50%
* 102013, Pt. A, Foil 1	/		/	/	/	/	/				60%
Foil 2								/			10%
Foil 3									/		10%
Foil 4										/	10%
(I Don't Know) Foil 5		/									10%
* 101003, Pt. A, 1 or More Correct Responses	/	/	/		/	/	/	/	/		80%
At Least One Category 10	/	/	/		/	/					50%
At Least One Category 11		/	/	/	/	/		/	/		70%
At Least One Category 12	/						/		/		30%

FIGURE 12 (cont.). Sample Tally Sheet

Exercises and Scoring Categories	John	Maria	Robert	Frank	Linda	Tina	Susan	Michael	Ann	Julius	Percentage for Correct Item Part
At Least One Category 20										/	20%
101003, Pt. A, At Least One Category 21			/								20%
102018, Pt. D, Foil 1		/						/			20%
* Foil 2	/		/				/			/	40%
Foil 3				/					/		20%
Foil 4					/	/					20%
(I Don't Know) Foil 5											0%
											Average % Correct on 4 Items, 62.5%
% Correct Each Student (Uses Only Asterisked Lines)	100%	50%	75%	50%	75%	75%	100%	25%	25%	50%	

**Indicates line is to be used for summary scores.*

sidered as a part of the total for the purpose of calculating percentages. These people should be tallied in a separate category, called something like "branched around" or "not applicable," and the total number of students taking the test should be used as the basis for calculating percentages for each part of the exercises.

Possible Modifications of Scoring Procedures

We have stressed the National Assessment procedures because they are necessary for comparing results and because they are helpful in understanding how

the NAEP materials were originally intended to be used. We hope you will feel free to experiment with other methods of administration, scoring, analysis and uses, as long as you do not then go back and try to compare your results with National Assessment results.

As noted previously, the most important thing to consider in designing a scoring system is that the system treat all student responses consistently. Teachers, schools and districts can devise their own methods of scoring and tallying results, perhaps using procedures already established for other testing activities. If you have modified the exercise slightly, the scoring guides should be reviewed and, if necessary, modified accordingly.

If you have written new exercises or made very extensive modifications to the items given, you may have to develop entirely new scoring guides. First, determine the categories of responses -- both correct and incorrect -- that are of interest to you. If you are using the items with a small group of students, you may be able to complete the scoring as you would with any teacher-constructed test. For a larger group, try out the new or modified items by administering them to a small sample of students. Score the responses using your preliminary scoring guides. You may discover that responses are given that do not fit your scoring guide categories, that the wording of the question elicits responses different from those expected and so on. At this point, make modifications to your items and scoring guides. When scoring the entire set of papers, you may find that further modifications, not called for by the small tryout sample of papers, are necessary. If you feel that the problem with the scoring guide is serious, it may be necessary to change the scoring guide and rescore the papers that have already been scored.

Possible Modifications of Exercise Tallying Procedures

Tallying of results will be highly dependent upon your intended use of the data. You may wish to change some of National Assessment's specifications for certain exercises.

Modifications to specific exercises might simply consist of changing the number of responses required for an acceptable response level. Another change might be to combine exercise parts, for example, asking that students answer all or some combination of exercise parts correctly to be scored correct on the exercises.

APPENDIX A

DISCUSSION ON SURVEY SAMPLING

The purpose of this discussion is to acquaint readers who are not familiar with sampling theory or techniques with some basic facts, limitations and applications of sampling so that they can decide whether to use these procedures for conducting an assessment using the career and occupational development kit. The first part of the discussion presents an overview of sampling, its benefits and limitations. This information should be helpful in guiding the reader's decision whether to take a sampling approach rather than assess all students. The rest of this discussion presents step-by-step instructions for two simplified sampling procedures. Neither approach requires extensive calculations, since information such as appropriate sample sizes and estimates of standard error is presented in tabular form. For readers with backgrounds in sampling statistics who wish to explore more elaborate sampling alternatives than will be discussed here, we refer them to two references that will be useful.[1,2]

Deciding Whether Sampling Can Be of Benefit to You

The techniques of sampling enable you to obtain reasonably accurate estimates of performance levels of groups of students without having to test each and every student. Sampling is not, however, without limitations. The most important limitation is that its use precludes your being able to identify performance levels of individual students who were not included in the sample. If you wish to have information about every student, sampling will not be appropriate.

A second limitation stems from the fact that sampling is most efficient, from a cost-savings point of view, in schools or school systems that have large enrollments. As a general rule, the larger the enrollment, the smaller the sample can be as a percentage of the total enrollment and still provide accurate information . With small enrollments, it may be necessary to assess almost all students in a particular age or grade to assure accurate estimates. To decide whether your school or district (for sampling purposes) falls into

---

[1]W.G. Cochran, Sampling Techniques (2nd ed.) (New York: Wiley, 1963).
[2]L. Kish, Survey Sampling (New York: Wiley, 1965).

either a "large enrollment category" or a "small enrollment category," you will need to relate your grade or age enrollment data to the information provided in Table A-1. To use this table, match the total number of students you are studying (grade level, age, etc.) with the number in the column labeled "total number of eligibles" that most closely approximates it. For example, if you wish to assess eighth graders and your eighth grade enrollment is 797, find the row for 800 eligibles. As shown by the numbers to the right, the sample sizes necessary to produce reliable data range from 740 students (very reliable) down to 270 students (less reliable). In this example, a sample size of 740 is estimated in Table A-1 to result in a standard error of 0.5 percentage points. Thus, if the percentage of students responding correctly on a given question is found to be 50%, we are 95% confident that the percentage for all eligibles would be 50% ± two standard errors. This provides a confidence interval of 49% to 51%. In contrast, a sample size of 270 is estimated to result in a standard error of 2.5 percentage points. With this size sample, a 95% confidence interval when p = 50% would be 45% to 55%.

The enrollment used in this example is large enough that sampling ought to be considered seriously as an option. Whether to assess 740 students rather than a minimal sample of 270 should be decided after taking into account the sampling error that the different sample sizes are likely to produce. This decision should be made by school personnel who are familiar with the intended uses of the assessment information and who have a feel for the level of accuracy needed.

In studying this table, it becomes evident that sampling is potentially most useful when numbers to be assessed exceed 400-500 students. With smaller numbers, it is necessary to test almost all students to obtain precise estimates from sampling. Even if greater error can be tolerated in the findings, one must assess nearly half of the students.

A final disadvantage of sampling to be discussed here stems from possible disruptions to ongoing classroom instruction. The step-by-step procedures to be discussed later in this appendix require that students be sampled randomly from a school-wide or district-wide listing by some criteria. Selection of students from such a listing means that at the time of testing, the students will need to leave their regularly assigned classes and come to a central site to participate.

Overview of Steps in Sampling

Sampling requires planning and procedural steps that are not needed when all students are assessed. Steps of both kinds are outlined below, and then are discussed in more detail.

Planning Steps Involved

1. Identify the population(s) of interest for purposes of comparing your data to National Assessment data.

TABLE A-1. Table of Approximate Sample Sizes for Various Levels of Sampling Error

Total Number of Eligibles	Estimated Size of One Standard Error*		
	0.5%	1.0%	2.5%
100	100	100	100
150	150	150	110
200	200	185	135
250	250	230	160
300	300	270	175
350	340	310	190
400	390	350	200
450	430	380	215
500	475	415	225
600	570	485	240
700	650	550	255
800	740	600	270
900	825	665	275
1000	910	715	285
1250	1110	835	300
1500	1300	940	315
1750	1500	1030	325
2000	1667	1110	335
2500	2000	1250	350
3000	2300	1365	355
4000	2850	1540	365

*The percentage (P) answering correctly $\pm$ 2 times standard error provides an interval that we are 95% confident would contain the percentage answering correctly of all age eligibles. If P = 50% and one standard error equals 1.0%, the 95% confidence interval is 48% to 52%.

2. Decide whether there are any subgroups of the above populations for which information is needed, such as males v. females, black students v. white students, students attending school A v. students attending school B, etc.

3. Decide the amount of error you are willing to tolerate in the findings (see Table A-1 for amount of estimated error associated with data from different sample sizes).

4. Decide which one of the two sampling approaches described later in this discussion will be more useful to you.

Procedural Steps Involved

1. Prepare a listing of eligible students. The list should be arranged alphabetically by student's last name. Other student information such as sex, race, birthdate or school attended may also be needed and will be discussed later.

2. Identify the number of students needed to be sampled in accord with the amount of error to be tolerated.

3. Select randomly the particular students who will compose the sample.

4. Notify students and have them answer the item.

5. Tally the data and compute the percent correct for each question based on sample data.

6. Determine the reliability of the comparisons between your data and those provided by National Assessment.

Choosing Which Sampling Approach Will Be Most Useful for Your Purposes

This manual is intended to help small- and medium-size districts who may not have measurement or statistical specialists on staff to conduct an assessment following National Assessment procedures. The two sampling procedures described are straightforward, easy to implement and, if done carefully, result in reliable information. The major difference in the two procedures is the level of information that they provide.

The first type of sampling is called simple random sampling. It is most appropriate when only district-wide or school-wide information is required and information on certain subgroups is not essential. For example, if your main interest is to answer the question "How well do students in my district or school compare as a total group to National Assessment findings?" simple random sampling can provide the answer.

On the other hand, if acquiring information on subgroups of students is highly desired or essential, then a procedure called stratified random sampling is recommended. For example, if your main purpose is to compare the relative performances of students in school A v. school B, stratified random sampling should be followed. This procedure will help assure that your sample will automatically include enough students from each school to make valid comparisons possible.

Simple Random Sampling

The word "random" may mean different things under different circumstances. As the word is used here, it means providing an equal opportunity for each eligible student to be included in the sample. Thus, it is necessary to prepare a listing of names of all eligible students before the selection can take place.

Prepare the listing. With simple random sampling, any ordering of student names is satisfactory but alphabetical ordering is frequently used. This may be done by clerks or in some schools can be provided by computer. In some instances it may be helpful to note additional student or school information alongside names when preparing this listing. If, for example, a specific time and date for the assessment has been set, it may be useful to list the class assignment of each student for that time and date so, if selected, the teacher can be notified. Or it may be easier to do this after the selection has been made inasmuch as this information will be needed only for the students actually sampled.

The next step is to number the list sequentially from 1 to N, where N is the total number of eligibles. No student's name should appear more than once, nore should any one number appear more than once.

Determine sample size. Table A-1 should be consulted to determine the sample size needed, taking into account the amount of error that is likely to result from samples of different sizes. Keep in mind that the numbers shown are approximate sample sizes that can be adjusted somewhat if necessary. For instance, if your auditorium holds 300 students and the proper sample size for your particular enrollment is 310, you may sample 300 only.

Randomly select students. An effective means of selecting students randomly is to use a table of random numbers found in many statistics textbooks or in mathematics reference books. Beginning with the first random number listed and taking additional numbers in sequence, select students from the listing sheet who have corresponding numbers. If you encounter a random number larger than any number on your listing sheet, disregard it. If you encounter the same number twice, also disregard it. Continue this process until the appropriate number of students have been selected. It is necessary to keep a tally of the number of students selected so that you know when the proper sample size has been reached.

Students so selected constitute the sample to be assessed. Every effort should be made to assess these students and only these students.

Stratified Random Sampling

Stratified random sampling provides two major advantages over simple random sampling. First, it can be used to help assure that certain subgroups will be adequately represented in the sample so that these subgroup comparisons can be made. For instance, if it is essential to be able to compare students in school A with students in school B, stratified sampling should be employed.

A second advantage is that smaller standard errors are often obtained, or smaller sample sizes can be used (as compared to simple random sampling) if stratified sampling is employed. One can achieve these benefits from this type of sample even though you do not intend to make comparisons among the stratification variables or subgroups.

Preparing the listing. The idea behind stratified sampling is to divide the list of eligibles into relatively homogeneous parts or strata such as separate schools, separate grades within schools or scholastic grade or previous test results -- "A" students, "B" students, etc. Stratification according to performance variables such as previous test results or student grade averages serves to improve the accuracy of estimates provided through sampling. Stratification according to reporting variables such as school attended may increase accuracy but is intended mainly to assure that you will be able to make comparisons between subgroups of students.

Whatever variables are used, eligibles should be classified separately at the time the student eligibility lists are prepared. For example, if your intention is to compare students from school A with students from school B, you would prepare two separate lists. Number student names sequentially beginning with 1 on the first list and continue the number sequence onto the second list as necessary.

Determine sample size. Refer to Table A-1 to determine the total number of students to be sampled, again taking into account the size of the standard error that is likely to result from samples of different sizes. If the number of eligibles on each listing is different, as is likely to be the case, it will be necessary to properly apportion the total sample size to each of the separate listings. This is accomplished by computing the sampling fraction (f) and using it to determine the number of students to be sampled from each listing.

$$f = \frac{n}{N}$$

where n = total sample size needed,

N = total number of eligibles.

For example, assume that the district-wide eighth grade enrollment is 1,000. From Table A-1 you decide to sample 715 (one standard error estimated to be 1%). The fraction (f) then is given by

$$f = \frac{715}{1,000}$$

$$= .715.$$

Further assume that you have two listings -- one for school A with 400 students and another for school B with 600 students (400 + 600 = 1,000). In this example, you would select 286 students (.715 X 400 = 286) from school A and 429 students (.715 X 600 = 429) from school B.

At this point it is necessary to note a possible limitation of Table A-1 in estimating sample sizes when stratified sampling is used. Since Table A-1 was designed to estimate sample sizes for the total population, its use will not always assure a sufficient sample size for certain subgroups.

For example, assume that your district enrollment for grade 8 again is 1,000 and you again decide to sample 715. However, now assume that 800 attend school A and only 200 attend school B.

After applying the sampling fraction (f) to the eligibility list for each school, you find that 572 are to be sampled from school A and 143 from school B. The problem that arises is that the disproportionate enrollments between the two schools result in only 143 students being assessed in school B. Table A-1 states that when the number of eligibles is only 200, it is necessary to sample 185 to assure the stated size of sampling error. Under the circumstances you have two choices:

1. Sample 572 from school A and 143 from school B with the recognition that data from school B probably will contain larger sampling errors than data from school A.

2. Sample 572 from school A and test all students in school B, thereby eliminating sampling errors for school B students.

Randomly select students. The procedure for selecting students is the same as that described in the preceding section, "Prepare the listing," except that two or more lists are being considered. As the random numbers are identified from the table, determine whether any of the student listings contain a corresponding number. If so, that student is selected. Upon reaching the proper sample size for any one list, continue this process for the remaining lists until the proper sample size has been reached for each list.

Be sure that these students and only these students so selected are assessed.

Determining Population Estimates From Sample Data

The intent of sampling, of course, is to estimate how well the total group has performed even though not all students were tested.

In simple random sampling and in stratified random sampling as described here, the sample data provides a direct estimate of the entire age or grade population. Thus, for each question, one need only compute the percent correct for the population. The first step is to find the proportion (p) of students in the sample responding correctly in which

$$p = \frac{n_c}{n} = \left(\frac{\text{number of students responding correctly}}{\text{number of students responding}}\right) .$$

The percent correct (P) is then found by multiplying p X 100 to give P. This process is repeated for each question. Since some students who were sampled may not take the survey, the number of students responding may not equal the sample size in all cases.

In addition to computing percent correct on total group performance for each question, you may wish to compute percent correct for certain subgroups such as males v. females. The formula for computing percent correct for subgroups is the same as used for the total group except only subgroup members are counted. Again the first step is to find the proportion (p) of sample students in the subgroup responding correctly in which

$$p = \frac{n_c}{n} = \frac{\text{number of students in subgroup responding correctly.}}{\text{number of students in subgroup responding}}$$

Thus, to compute proportion correct for males and females,

$$p_m = \frac{n_c}{n} = \frac{\text{number of males responding correctly}}{\text{number of males responding}}$$

and

$$p_f = \frac{n_c}{n} = \frac{\text{number of females responding correctly.}}{\text{number of females responding}}$$

The percent correct (P) for each group is then found by multiplying each p X 100.

The reliability of subgroup data depends partly on the number of respondents within those subgroups that happen to have been selected in the sample. Males and females, as a subgroup, probably can be broken out since approximately half of the sample will be comprised of males and half of females. For other subgroups such as white or black students, it may be the case that too few whites or blacks were selected in the sample to result in reliable esti-

mates. Procedures for judging the reliability of the findings will be discussed next.

Estimating the Reliability of Your Findings

After you have completed the computations, question by question, for determining the percent correct for total group and for subgroups of interest, you may also wish to determine how reliable these data are. For instance, how sure can you be that the sample data accurately estimates the findings you might obtain had all students been assessed? Or, how sure can you be that subgroup data are really accurate estimates?

To assist in making these determinations, we have provided estimates of standard error that usually result from samples of various sizes in Tables A-2 through A-9. Error estimates in these tables tend to be larger than those you might obtain if you actually calculate the error in your particular sample according to standard statistical formulas. In other words, you can be reasonably sure that standard errors that you might compute from your sample will not be larger than those estimated by the tables.

In order to use the tables you will need to provide three kinds of information:

1. Your sample size. You must know the number of students in the total sample and the number of students in the sample making up particular subgroups such as the number of males and the number of females.

2. Percent correct. You must know for each question the percent correct for the total sample and the percent correct for any subgroups of interest within the sample.

3. Sampling fraction. You must compute the sampling fraction (f) for the total sample (see formula under preceding section, "Determine sample size"). Depending upon the value of f you will need to refer to different tables. The tables cover values of f from 0.0 to .79.

The following example will explain how to use Tables A-2 through A-9. Suppose that there are 1,000 eligibles in your district and you decided from Table A-1 to sample 715 students because that size sample was estimated (by Table A-1) to produce sampling errors of about 1%. Suppose further that you stratified by school according to the sampling fraction .715 so that 286 of 400 eligible students were sampled from school A and 429 of 600 eligible students were sampled from school B. After sampling, you counted the number of males and females in the sample and found 370 males and 345 females.

For question one, the percent correct for each group was found to be as follows:

$$P_{total} = 64\%$$

$$P_{\text{school A}} = 83\%$$

$$P_{\text{school B}} = 51\%$$

$$P_{\text{males}} = 57\%$$

$$P_{\text{females}} = 71\%$$

We will first determine the standard error for the total group. Since different tables provide data for different ranges of f, we go to Table A-9 because this table is for f = .70 - .79; and f in this example is .715. Next we locate the sample size that is closest to 715. This is found to be 700. We then locate the percent correct that is closest to 64%. At the intersection of the row showing sample size and the column showing percent correct is a number, .90, which is the standard error (se) given in percentage points. The percent correct, plus and minus two times this standard error provides an interval that we are 95% confident contains the true population value. In this case, this interval is 62.2% - 65.8%.

The next example shows how to determine the reliability of male and female data. Using Table A-9 for a sample size of 370 (350) and for percent correct of 57% (55%), we find the standard error for males to be 1.33 percentage points.

With a sample size of 345 (350) and percent correct of 71% (70), Table A-9 gives 1.23 as the standard error for females.

The question of interest now is whether the findings for males and females are statistically significant. To test this we first determine the difference value (D) as follows:

$$D = P_m - P_f$$

$$= 57\% - 71\%$$

$$= -14\%.$$

Next we compute the standard error of the difference as follows:[3]

$$se_D = \sqrt{se_{Pm}^2 + se_{pf}^2}$$

---

[3]Many hand calculators compute square roots directly.

$$= \sqrt{(1.33)^2 + (1.23)^2}$$

$$= \sqrt{3.2818}$$

$$= 1.81$$

If $\frac{D}{se_D}$

is larger than plus or minus 2.0, then the difference (D) is significant at the 95% level of confidence. In this example, since

$$\frac{D}{se_D} = \frac{-14\%}{1.81} = -7.73$$

the difference between male and female findings is statistically significant.

In a final example comparing school A and school B, we will show only the results so that you can verify your own computations.

$$D = P_A - P_B$$

$$= 32$$

$$se_D = \sqrt{se_{P_A}^2 + se_{P_B}^2}$$

$$= 1.57\%$$

$$\frac{D}{se_D} = 20.38.$$

Comparing Your Data to National Assessment Data

Comparisons to National Assessment reporting groups (defined in Appendix C) will be limited by the number of students of a particular group that actually show up in your sample. National Assessment gathers information nationwide in all types of schools, but your sample (or even your population) may not contain enough students of a particular type to make certain comparisons meaningful. For instance, you may have very few students whose parents have not completed high school. Your sample may not contain any students like this. Thus, it would be meaningless to compare your findings to the particular category of NAEP findings that includes only students of this kind.

The procedures for making comparisons to National Assessment data are very similar to the ones described previously in the section on "Estimating the Reliability of Your Findings." The major difference is that NAEP data are used for percent correct and for standard errors along with your data in the formulas.

To compare your percent correct to National Assessment percent correct for each question compute:

$$D = P_{yours} - P_{NAEP}$$

and

$$se_D = \sqrt{se_{P_{yours}}^2 + se_{P_{NAEP}}^2}$$

and determine if $\frac{D}{se_D}$ is larger than ± 2.0.

To compare any subgroup such as males compute:

$$D_{males} = P_{your_{males}} - P_{NAEP_{males}}$$

and

$$se_{D_{males}} = \sqrt{se_{P_{your\ males}}^2 + se_{P_{NAEP\ males}}^2}$$

and determine if $\frac{D_{males}}{se_{D_{males}}}$ is larger than ± 2.0.

TABLE A-2. Estimates of Standard Errors for P-Values When Sampling Fractions Are

[redacted] 0 to .09

If the sample size n=[redacted] and the percent correct (p-value) is computed to be [redacted]

5% or 95%	10% or 90%	15% or 85%	20% or 80%	25% or 75%	30% or 70%	35% or 65%	40% or 60%	45% or 55%	50%
one standard error equals:									
n									
[redacted]									
2.13	2.94	3.50	3.92	4.24	4.49	4.67	4.80	4.87	4.90
1.74	2.40	2.85	3.19	3.46	3.66	3.81	3.91	3.97	3.99
1.51	2.07	2.47	2.76	2.99	3.17	3.30	3.38	3.44	3.45
1.35	1.85	2.21	2.47	2.67	2.83	2.95	3.03	3.07	3.09
1.23	1.69	2.01	2.25	2.44	2.58	2.69	2.76	2.80	2.82
1.14	1.57	1.86	2.09	2.26	2.39	2.49	2.56	2.60	2.61
1.06	1.46	1.74	1.95	2.11	2.24	2.33	2.39	2.43	2.44
1.00	1.38	1.64	1.84	1.99	2.11	2.19	2.25	2.29	2.30
0.95	1.31	1.56	1.75	1.89	2.00	2.08	2.14	2.17	2.18
0.87	1.19	1.42	1.59	1.72	1.82	1.90	1.95	1.98	1.99
0.80	1.11	1.32	1.47	1.60	1.69	1.76	1.81	1.83	1.84
0.75	1.03	1.23	1.38	1.49	1.58	1.64	1.69	1.72	1.72
0.71	0.98	1.16	1.30	1.41	1.49	1.55	1.59	1.62	1.63
0.67	0.93	1.10	1.23	1.34	1.41	1.47	1.51	1.53	1.54
0.55	0.76	0.90	1.01	1.09	1.15	1.20	1.23	1.25	1.26
0.48	0.65	0.78	0.87	0.94	1.00	1.04	1.07	1.08	1.09
0.42	0.58	0.70	0.78	0.84	0.89	0.93	0.96	0.97	0.97
0.39	0.53	0.64	0.71	0.77	0.82	0.85	0.87	0.89	0.89
0.36	0.49	0.59	0.66	0.71	0.76	0.79	0.81	0.82	0.82
0.34	0.46	0.55	0.62	0.67	0.71	0.74	0.76	0.77	0.77

TABLE A-3. Estimates of Standard Errors for P-Values When Sampling Fractions Are

.10 to .19

If the sample size n= [redacted] and the percent correct (p-value) is computed to be [redacted]

5% or 95%	10% or 90%	15% or 85%	20% or 80%	25% or 75%	30% or 70%	35% or 65%	40% or 60%	45% or 55%	50%
one standard error equals:									
n									
[redacted]	2.02	2.78	3.31	3.71	4.01	4.25	4.42	4.54	4.61	4.63
[redacted]	1.65	2.27	2.70	3.02	3.27	3.46	3.60	3.70	3.76	3.78
[redacted]	1.42	1.96	2.33	2.61	2.83	2.99	3.12	3.20	3.25	3.27
[redacted]	1.27	1.75	2.09	2.34	2.53	2.68	2.79	2.86	2.91	2.92
[redacted]	1.16	1.60	1.90	2.13	2.31	2.44	2.54	2.61	2.65	2.67
[redacted]	1.08	1.48	1.76	1.97	2.14	2.26	2.35	2.42	2.46	2.47
[redacted]	1.01	1.38	1.65	1.85	2.00	2.12	2.20	2.26	2.30	2.31
[redacted]	0.95	1.31	1.55	1.74	1.88	1.99	2.08	2.13	2.16	2.18
[redacted]	0.90	1.24	1.47	1.65	1.79	1.89	1.97	2.02	2.05	2.06
[redacted]	0.82	1.13	1.35	1.51	1.63	1.73	1.80	1.85	1.87	1.88
[redacted]	0.76	1.05	1.25	1.39	1.51	1.60	1.66	1.71	1.73	1.74
[redacted]	0.71	0.98	1.16	1.30	1.41	1.49	1.56	1.60	1.62	1.63
[redacted]	0.67	0.92	1.10	1.23	1.33	1.41	1.47	1.51	1.53	1.54
[redacted]	0.64	0.88	1.04	1.17	1.26	1.34	1.39	1.43	1.45	1.46
[redacted]	0.52	0.71	0.85	0.95	1.03	1.09	1.14	1.17	1.18	1.19
[redacted]	0.45	0.62	0.74	0.82	0.89	0.94	0.98	1.01	1.03	1.03
[redacted]	0.40	0.55	0.66	0.74	0.80	0.85	0.88	0.90	0.92	0.92
[redacted]	0.37	0.51	0.60	0.67	0.73	0.77	0.80	0.82	0.84	0.84
[redacted]	0.34	0.47	0.56	0.62	0.67	0.71	0.74	0.76	0.78	0.78
[redacted]	0.32	0.44	0.52	0.58	0.63	0.67	0.70	0.71	0.73	0.73

TABLE A-4. Estimates of Standard Errors for P-Values When Sampling Fractions Are

.20 to .29

If the sample size n= [redacted] and the percent correct (p-value) is computed to be [redacted]

n	5% or 95%	10% or 90%	15% or 85%	20% or 80%	25% or 75%	30% or 70%	35% or 65%	40% or 60%	45% or 55%	50%
[redacted]	1.90	[illegible]	3.11	3.48	3.77	3.99	4.15	4.26	4.33	4.35
[redacted]	1.55	2.13	2.53	2.84	3.07	3.25	3.38	3.48	3.53	3.55
[redacted]	1.34	1.84	2.19	2.46	2.66	2.81	2.93	3.01	3.05	3.07
[redacted]	1.20	1.65	1.96	2.20	2.38	2.52	2.62	2.69	2.73	2.74
[redacted]	1.09	1.50	1.79	2.00	2.17	2.30	2.39	2.45	2.49	2.50
[redacted]	1.01	1.39	1.66	1.85	2.01	2.12	2.21	2.27	2.31	2.32
[redacted]	0.94	1.30	1.55	1.73	1.88	1.99	2.07	2.12	2.16	2.17
[redacted]	0.89	1.23	1.46	1.63	1.77	1.87	1.95	2.00	2.03	2.04
[redacted]	0.84	1.16	1.38	1.55	1.68	1.78	1.85	1.90	1.93	1.94
[redacted]	0.77	1.06	1.26	1.42	1.53	1.62	1.69	1.73	1.76	1.77
[redacted]	0.71	0.98	1.17	1.31	1.42	1.50	1.56	1.60	1.63	1.64
[redacted]	0.67	0.92	1.09	1.23	1.33	1.40	1.46	1.50	1.52	1.53
[redacted]	0.63	0.87	1.03	1.16	1.25	1.32	1.38	1.41	1.44	1.44
[redacted]	0.60	0.82	0.98	1.10	1.19	1.26	1.31	1.34	1.36	1.37
[redacted]	0.49	0.67	0.80	0.89	0.97	1.03	1.07	1.10	1.11	1.12
[redacted]	0.42	0.58	0.69	0.77	0.84	0.89	0.92	0.95	0.96	0.97
[redacted]	0.38	0.52	0.62	0.69	0.75	0.79	0.83	0.85	0.86	0.87
[redacted]	0.34	0.47	0.56	0.63	0.68	0.72	0.75	0.77	0.79	0.79
[redacted]	0.32	0.44	0.52	0.59	0.63	0.67	0.70	0.72	0.73	0.73
[redacted]	0.30	0.41	0.49	0.55	0.59	0.63	0.65	0.67	0.68	0.68

TABLE A-5. Estimates of Standard Errors for P-Values When Sampling Fractions Are

.30 to .39

If the sample size n= [redacted] and the percent correct (p-value) is computed to be [redacted]

5% or 95%	10% or 90%	15% or 85%	20% or 80%	25% or 75%	30% or 70%	35% or 65%	40% or 60%	45% or 55%	50%
one standard error equals:									
n									
[redacted]	1.77	2.43	2.89	3.24	3.51	3.71	3.86	3.97	4.03	4.05
	1.44	1.98	2.36	2.64	2.86	3.03	3.15	3.24	3.29	3.30
	1.25	1.71	2.04	2.29	2.47	2.62	2.73	2.80	2.84	2.86
	1.11	1.53	1.82	2.04	2.21	2.34	2.44	2.50	2.54	2.55
	1.02	1.40	1.66	1.87	2.02	2.14	2.22	2.28	2.32	2.33
	0.94	1.29	1.54	1.73	1.87	1.98	2.06	2.11	2.15	2.16
	0.88	1.21	1.44	1.61	1.75	1.85	1.93	1.98	2.01	2.02
	0.83	1.14	1.36	1.52	1.65	1.74	1.81	1.86	1.89	1.90
	0.79	1.08	1.29	1.44	1.56	1.65	1.72	1.77	1.80	1.80
	0.72	0.99	1.18	1.32	1.43	1.51	1.57	1.61	1.64	1.65
	0.66	0.91	1.09	1.22	1.32	1.40	1.45	1.49	1.52	1.52
	0.62	0.86	1.02	1.14	1.24	1.31	1.36	1.40	1.42	1.43
	0.59	0.81	0.96	1.08	1.16	1.23	1.28	1.32	1.34	1.34
	0.56	0.77	0.91	1.02	1.10	1.17	1.22	1.25	1.27	1.28
	0.45	0.62	0.74	0.83	0.90	0.95	0.99	1.02	1.04	1.04
	0.39	0.54	0.64	0.72	0.78	0.83	0.86	0.88	0.90	0.90
	0.35	0.48	0.58	0.65	0.70	0.74	0.77	0.79	0.80	0.81
	0.32	0.44	0.53	0.59	0.64	0.67	0.70	0.72	0.73	0.74
	0.30	0.41	0.49	0.55	0.59	0.62	0.65	0.67	0.68	0.68
	0.28	0.38	0.46	0.51	0.55	0.58	0.61	0.62	0.63	0.64

TABLE A-6. Estimates of Standard Errors for P-Values When Sampling Fractions Are

.40 to .49

If the sample size n= [redacted] and the percent correct (p-value) is computed to be [redacted]

n	5% or 95%	10% or 90%	15% or 85%	20% or 80%	25% or 75%	30% or 70%	35% or 65%	40% or 60%	45% or 55%	50%
	one standard error equals:									
[redacted]	1.62	2.24	2.66	2.98	3.23	3.42	3.56	3.65	3.71	3.73
	1.32	1.82	2.17	2.43	2.63	2.78	2.90	2.98	3.02	3.04
	1.15	1.58	1.88	2.10	2.28	2.41	2.51	2.58	2.62	2.63
	1.02	1.41	1.68	1.88	2.04	2.15	2.24	2.30	2.34	2.35
	0.93	1.29	1.53	1.72	1.86	1.97	2.05	2.10	2.13	2.14
	0.87	1.19	1.42	1.59	1.72	1.82	1.89	1.94	1.97	1.98
	0.81	1.11	1.33	1.49	1.61	1.70	1.77	1.82	1.85	1.86
	0.76	1.05	1.25	1.40	1.52	1.60	1.67	1.71	1.74	1.75
	0.72	1.00	1.19	1.33	1.44	1.52	1.58	1.63	1.65	1.66
	0.66	0.91	1.08	1.21	1.31	1.39	1.45	1.48	1.51	1.52
	0.61	0.84	1.00	1.12	1.21	1.29	1.34	1.37	1.40	1.40
	0.57	0.79	0.94	1.05	1.14	1.20	1.25	1.29	1.31	1.31
	0.54	0.74	0.88	0.99	1.07	1.13	1.18	1.21	1.23	1.24
	0.51	0.70	0.84	0.94	1.02	1.08	1.12	1.15	1.17	1.17
	0.42	0.57	0.68	0.77	0.83	0.88	0.91	0.94	0.95	0.96
	0.36	0.50	0.59	0.66	0.72	0.76	0.79	0.81	0.83	0.83
	0.32	0.45	0.53	0.59	0.64	0.68	0.71	0.73	0.74	0.74
	0.30	0.41	0.48	0.54	0.59	0.62	0.65	0.66	0.67	0.68
	0.27	0.38	0.45	0.50	0.54	0.57	0.60	0.61	0.62	0.63
	0.26	0.35	0.42	0.47	0.51	0.54	0.56	0.57	0.58	0.59

TABLE A-7. Estimates of Standard Errors for P-Values When Sampling Fractions Are

.50 to .59

If the sample size n= [redacted] and the percent correct (p-value) is computed to be [redacted]

5% or 95%	10% or 90%	15% or 85%	20% or 80%	25% or 75%	30% or 70%	35% or 65%	40% or 60%	45% or 55%	50%
one standard error equals									
n									
[redacted]	1.47	2.02	2.41	2.70	2.92	3.09	3.22	3.30	3.35	3.37
[redacted]	1.20	1.65	1.96	2.20	2.38	2.52	2.62	2.69	2.73	2.75
[redacted]	1.04	1.43	1.70	1.90	2.06	2.18	2.27	2.33	2.37	2.38
[redacted]	0.93	1.28	1.52	1.70	1.84	1.95	2.03	2.08	2.11	2.13
[redacted]	0.85	1.16	1.39	1.55	1.68	1.78	1.85	1.90	1.93	1.94
[redacted]	0.78	1.08	1.28	1.44	1.55	1.65	1.71	1.76	1.79	1.80
[redacted]	0.73	1.01	1.20	1.34	1.45	1.54	1.60	1.65	1.67	1.68
[redacted]	0.69	0.95	1.13	1.27	1.37	1.45	1.51	1.55	1.57	1.58
[redacted]	0.65	0.90	1.07	1.20	1.30	1.38	1.43	1.47	1.49	1.50
[redacted]	0.60	0.82	0.98	1.10	1.19	1.26	1.31	1.34	1.36	1.37
[redacted]	0.55	0.76	0.91	1.01	1.10	1.16	1.21	1.24	1.26	1.27
[redacted]	0.52	0.71	0.85	0.95	1.03	1.09	1.13	1.16	1.18	1.19
[redacted]	0.49	0.67	0.80	0.89	0.97	1.03	1.07	1.10	1.11	1.12
[redacted]	0.46	0.64	0.76	0.85	0.92	0.97	1.01	1.04	1.06	1.06
[redacted]	0.38	0.52	0.62	0.69	0.75	0.79	0.83	0.85	0.86	0.87
[redacted]	0.33	0.45	0.54	0.60	0.65	0.69	0.72	0.74	0.75	0.75
[redacted]	0.29	0.40	0.48	0.54	0.58	0.61	0.64	0.66	0.67	0.67
[redacted]	0.27	0.37	0.44	0.49	0.53	0.56	0.58	0.60	0.61	0.61
[redacted]	0.25	0.34	0.40	0.45	0.49	0.52	0.54	0.56	0.56	0.57
[redacted]	0.23	0.32	0.38	0.42	0.46	0.49	0.51	0.52	0.53	0.53

TABLE A-8. Estimates of Standard Errors for P-Values When Sampling Fractions Are

.60 to .69

If the sample size n= [redacted] and the percent correct (p-value) is computed to be [redacted]

n	5% or 95%	10% or 90%	15% or 85%	20% or 80%	25% or 75%	30% or 70%	35% or 65%	40% or 60%	45% or 55%	50%
	one standard error equals:									
[redacted]	1.30	1.78	2.12	2.38	2.57	2.72	2.84	2.91	2.96	2.97
	1.06	1.45	1.73	1.94	2.10	2.22	2.31	2.37	2.41	2.42
	0.91	1.26	1.50	1.68	1.82	1.92	2.00	2.05	2.09	2.10
	0.82	1.12	1.34	1.50	1.62	1.72	1.79	1.84	1.87	1.87
	0.75	1.03	1.22	1.37	1.48	1.57	1.63	1.68	1.70	1.71
	0.69	0.95	1.13	1.27	1.37	1.45	1.51	1.55	1.58	1.58
	0.65	0.89	1.06	1.18	1.28	1.36	1.41	1.45	1.47	1.48
	0.61	0.84	1.00	1.12	1.21	1.28	1.33	1.37	1.39	1.40
	0.58	0.79	0.95	1.06	1.15	1.21	1.26	1.30	1.32	1.32
	0.53	0.73	0.86	0.97	1.05	1.11	1.15	1.18	1.20	1.21
	0.49	0.67	0.80	0.90	0.97	1.03	1.07	1.10	1.11	1.12
	0.46	0.63	0.75	0.84	0.91	0.96	1.00	1.03	1.04	1.05
	0.43	0.59	0.70	0.79	0.85	0.90	0.94	0.97	0.98	0.99
	0.41	0.56	0.67	0.75	0.81	0.86	0.89	0.92	0.93	0.94
	0.33	0.46	0.55	0.61	0.66	0.70	0.73	0.75	0.76	0.76
	0.29	0.40	0.47	0.53	0.57	0.61	0.63	0.65	0.66	0.66
	0.26	0.36	0.42	0.47	0.51	0.54	0.56	0.58	0.59	0.59
	0.24	0.32	0.39	0.43	0.47	0.50	0.52	0.53	0.54	0.54
	0.22	0.30	0.36	0.40	0.43	0.46	0.48	0.49	0.50	0.50
	0.20	0.28	0.33	0.37	0.41	0.43	0.45	0.46	0.47	0.47

TABLE A-9. Estimates of Standard Errors for P-Values When Sampling Fractions Are

.70 to .79

If the sample size n= [redacted] and the percent correct (p-value) is computed to be [redacted]

n	5% or 95%	10% or 90%	15% or 85%	20% or 80%	25% or 75%	30% or 70%	35% or 65%	40% or 60%	45% or 55%	50%
	one standard error equals:									
[redacted]	1.10	1.51	1.79	2.01	2.18	2.30	2.40	2.46	2.50	2.51
	0.89	1.23	1.46	1.64	1.77	1.88	1.95	2.01	2.04	2.05
	0.77	1.06	1.27	1.42	1.53	1.62	1.69	1.74	1.76	1.77
	0.69	0.95	1.13	1.27	1.37	1.45	1.51	1.55	1.58	1.58
	0.63	0.87	1.03	1.16	1.25	1.33	1.38	1.42	1.44	1.45
	0.58	0.80	0.96	1.07	1.16	1.23	1.28	1.31	1.33	1.34
	0.55	0.75	0.89	1.00	1.08	1.15	1.19	1.23	1.25	1.25
	0.51	0.71	0.84	0.94	1.02	1.08	1.13	1.16	1.17	1.18
	0.49	0.67	0.80	0.90	0.97	1.03	1.07	1.10	1.11	1.12
	0.45	0.61	0.73	0.82	0.88	0.94	0.97	1.00	1.02	1.02
	0.41	0.57	0.68	0.76	0.82	0.87	0.90	0.93	0.94	0.95
	0.39	0.53	0.63	0.71	0.77	0.81	0.84	0.87	0.88	0.88
	0.36	0.50	0.60	0.67	0.72	0.76	0.80	0.82	0.83	0.83
	0.34	0.47	0.56	0.63	0.68	0.72	0.75	0.77	0.79	0.79
	0.28	0.39	0.46	0.52	0.56	0.59	0.62	0.63	0.64	0.65
	0.24	0.34	0.40	0.45	0.48	0.51	0.53	0.55	0.56	0.56
	0.22	0.30	0.36	0.40	0.43	0.46	0.48	0.49	0.50	0.50
	0.20	0.27	0.33	0.37	0.40	0.42	0.44	0.45	0.45	0.46
	0.18	0.25	0.30	0.34	0.37	0.39	0.40	0.41	0.42	0.42
	0.17	0.24	0.28	0.32	0.34	0.36	0.38	0.39	0.39	0.40

APPENDIX B

TABLE OF ITEMS: SHOWING ADMINISTRATION TYPE, SCORING TYPE AND TIMINGS

Exercise and Age Levels	Exercise Part* Nos.	Administration Type†	Scoring Type**	Timing†† 9	13	17
101002-17		I	OE	---	---	---
101003-9		I	OE	---	---	---
101003-13	Part A	G	OE	---	62	---
	Part B	G	OE	---	62	---
101005-9,13,17		I	OE	---	---	---
101016-17		I	OE	---	---	---
102002-9,13	Part A	G	MC	51	49	---
	Part B	G	MC	45	44	---
	Part C	G	MC	41	42	---
	Part D	G	MC	44	43	---
	Part E	G	MC	44	43	---
102013-13,17	Part A	G	MC	---	33	27
	Part B	G	MC	---	27	33
	Part C	G	MC	---	26	21
	Part D	G	MC	---	26	21
	Part E	G	MC	---	26	23
	Part F	G	MC	---	28	23
	Part G	G	MC	---	27	23
	Part H	G	MC	---	29	24
	Part I	G	MC	---	30	26
102018-17	Part A	G	MC	---	---	32
	Part B	G	MC	---	---	28
	Part C	G	MC	---	---	27
	Part D	G	MC	---	---	26
	Part E	G	MC	---	---	30
102023-17	Part A	G	OE	---	---	114
	Part B	G	OE	---	---	115
102033-9		I	OE	---	---	---

*Parts are not listed for individual exercises since parts were not timed.
†G = group administered; I = individually administered.
**MC - multiple-choice; OE = open-ended.
††Timings are given in seconds.

Exercise and Age Levels	Exercise Part Nos.	Administration Type	Scoring Type	Timing 9	13	17
102036-13,17	Part A	G	MC	---	23	26
	Part B	G	MC	---	20	22
	Part C	G	MC	---	21	23
	Part D	G	MC	---	20	21
	Part E	G	MC	---	23	24
103002-13,17		I	OE	---	---	---
104003-13,17		I	OE	---	---	---
104009-13,17		G	OE	---	195	253
104011-13,17		G	OE	---	120	121
104015-13,17		I	OE	---	---	---
200001-17		I	OE	---	---	---
200016-9,13,17		G	OE	91	90	90
200018-9,13		G	MC	148	152	---
	Parts A-K			approx. 15 sec. ea.		
200019-9,13		G	MC	453	441	---
	Parts A-Z			approx. 15-20 sec. ea.		
301019		G	MC	---	59	60
301032		G	OE	---	203	143
302005		G	OE	---	497	353
302010		G	OE	134	134	130
302015		G	OE	366	249	---
304001-17		G	OE	---	---	204
305008-9,13,17		I	OE	---	---	---
306006-13,17		G	OE	---	135	136
306009-17	Choosing Ad	G	MC	---	---	125
	Writing Letter	G	OE	---	---	674
401003-9,13		G	MC	119	92	---
401009-13,17	Introduction	G	--	---	31	33
	Part A	G	MC	---	48	49
	Part B	G	OE	---	93	93
401013-9,13		I	OE	---	---	---
401013-17		G	OE	---	---	62
403005-9,13		G	OE	91	60	---
403010-17		G	MC	---	---	60
501009-13,17		G	OE	---	165	121
501011-17	Part A	G	MC	---	---	38
	Part B	G	OE	---	---	50
503011-9,13,17		G	OE	60	47	46
503012-9,13	Part A	G	MC	19	19	---
	Part B	G	MC	14	15	---
	Part C	G	MC	15	15	---
	Part D	G	MC	15	15	---
	Part E	G	MC	15	17	---
504005-13,17	Part A	G	MC	---	35	35
	Part B	G	OE	---	85	85

APPENDIX C

DEFINITIONS OF NATIONAL ASSESSMENT REPORTING GROUPS

Population Groups Included in the Study

Sex

Results are reported separately for males and females.

Race

Results are reported for blacks and whites.

Size and Type of Community

The groups within this variable are defined by the size of the community in which a respondent lives and an occupational profile of the area. All population sizes are based on data from the U.S. Bureau of the Census (1970). The following groups are reported:

Rural. These are areas where a high proportion of the residents are farmers or farm workers. At least some of the area is open country or contains towns of less than 2,500 population; no towns in the area are greater than 10,000 or are suburbs of medium or large cities.

Low metro (low-socioeconomic or improverished urban). Areas in or around cities with a population greater than 200,000 where a high proportion of the residents are on welfare or are not regularly employed.

High metro (high-socioeconomic or affluent urban and suburban communities). Areas in or around cities with a population greater than 200,000 where a high proportion of the residents are in professional or managerial positions.

Main big city. Respondents in this group live in big cities greater than 200,000 and are not included in either the low-metro or high-metro groups.

Urban fringe. These respondents live in the metropolitan areas of big cities greater than 200,000 and are not included in either the low-metro or high-metro groups.

Medium city. These respondents live in cities with populations between 25,000 and 200,000 that are not urbanized areas of big cities.

Small places. Respondents in this group live in open country or in places with populations of less than 25,000, excluding those in the rural group.

Region

The nation has been divided into four regions -- Southeast, Northeast, Central and West -- as shown in Exhibit 1.

EXHIBIT 1. National Assessment Geographic Regions

Parental Education

These groups are classified by the highest level of education attained by either parent, from no high school through post high school. The respondent reported how much school both his parents completed; the highest level attained by either parent is reported. The following groups are reported:

No high school. Neither parent completed schooling beyond the eighth grade.

Some high school. At least one parent completed some schooling in grades 9 through 12; neither completed high school.

Graduated high school. At least one parent graduated from high school; neither participated in post-high-school education.

Post high school. At least one parent had some education beyond high school.

School Grade

The school grades for 9-year-olds, 13-year-olds and in-school 17-year-olds are reported. At each age, grade is reported for levels one-four; the definitions of these levels are:

Grade Level	Age 9	Age 13	Age 17
1	Below 3rd grade	Below 7th grade	Below 10th grade
2	Grade 3	Grade 7	Grade 10
3	Grade 4	Grade 8	Grade 11
4	Above 4th grade	Above 8th grade	Grade 12

Home Environment

Respondents at all ages were asked the following questions:

1. Does your family get a newspaper?
2. Does your family get any magazines regularly?
3. Are there more than 25 books in your home?
4. Is there an encyclopedia in your home?

The number of these things in the respondent's home is reported as follows:

1. Less than three home-environment items.
2. Three home-environment items.
3. Four home-environment items.

APPENDIX D

NATIONAL ASSESSMENT SCORING RULES FOR OPEN-ENDED ITEMS

General Rules

1. When anything is erased, completely ignore it.

2. Ignore spelling and punctuation errors and bad grammar.

3. If an answer line is provided, what is written on the line takes precedence over writing elsewhere on the page. However, if the answer line is blank, categorize whatever is written on the page.

Definitions of Categories

1. No response -- should be used when the answer line is completely blank or when "that's all" or "no more" is written following a series of answers.

2. "I don't know" -- also include "can't tell," "beats me," "don't have any idea," "can't think of any more," "that's all I know." Giving a reason for not knowing is counted as an "I don't know" -- for example, "don't understand the question," "don't know the meaning of a word in the question," "I don't know because I've never thought much about it."

   If an answer is given, the "I don't know" category should not be used. Answers, whether right or wrong, take precedence over "I don't know."

3. Garbage -- includes such things as:

   a. Fragment of an answer -- "be," "because the person wants."

   b. Illegible -- a critical word/words cannot be made out.

   c. Illiterate or nonsensical -- answer does not make any sense in and of itself, not necessarily in relation to the question. (Note: This does not include people who misinterpret the question.)

Questions With Several Answers

1. Sometimes an answer goes over more than one answer line. Consider it as if it were all on the same line.

2. If more than one answer is given on a line and there are blank lines or I don't know answers, separate the answers and move the remaining answers down one line.

EXAMPLE:		SCORE:	
1.	A & B	1.	A
2.	C	2.	B
3.	D	3.	C
4.		4.	D
5.		5.	(NR)

   The two answers must be clearly two, and not both part of one idea. Whenever they fit in different categories, they are always separated when "possible"; when the category is the same, be reluctant to separate them.

3. If the answer cannot be separated or if the question has only one response line, the following rules apply:

   a. If both answers are acceptable, code the first answer unless the key indicates that one has precedence.

   b. If both are unacceptable, a specific category has precedence over "other unacceptable" and over nonsense "garbage" categories.

   c. If one answer is acceptable and one not, code the acceptable one.

   d. If three or more answers are given on a line and over half is incorrect, do not count the answer correct. In other words, the student must be 50% correct to be given credit.

Use of the "Exact Duplicates" Category

1. Be reluctant to code "exact duplicates." A general plus a specific both get credit if the general code includes more than the specific.

   Examples that would <u>not</u> be categorized as "exact duplicates":

   a. What are my duties?
      Will I be ringing up sales?

   b. Have a good education.
      Have a high school diploma.

   c. What is the pay?
      How often do I get paid?

d. Clean my bedroom.
   Clean the house.

Examples that would be categorized as "exact duplicates":

a. What are the hours of the job?
   What time do I come in? (exact duplicate)
   What time do I leave? (exact duplicate)

b. If you can handle the job.
   If you have the ability to do the job. (exact duplicate)